aws

INNOVATE

DATA EDITION

19 August 2021

Modernize log analytics with Amazon OpenSearch Service

Muhammad Ali

Sr. Analytics Solutions Architect
Amazon Web Services

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

- Amazon OpenSearch Service Introduction
- Why do you need Amazon OpenSearch Service?
- How to get started?
- Best Practices
- Security
- Demo

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why log analytics?

Machine generated data

- IT & DevOps
- Applications & Cloud infrastructure
- IoT & Wireless

Valuable Insights

- Systems insights
- Products insights
- User behaviors
- Security threat detection
- Anomalous behaviors

Right tool

- Manual text analysis is difficult
- Traditional databases do not scale well
- Data warehouse do not provide indexes

© 2020, Amazon Web Services, Inc. or its Affiliates.

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon OpenSearch Service

Amazon OpenSearch Service is a fully managed service that makes it easy to deploy, manage, and run OpenSearch cost effectively with industry-leading reliability, scalability, and security

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Benefits of Amazon OpenSearch Service

Fully managed

Get a production-ready cluster up and running in minutes; no more patching, versioning, and backups

Access to all data

Capture, retain, correlate, and analyze *all* data

Highly scalable and available

Resize your cluster with a few clicks or a single API call; replicate data across multiple Availability Zones

Secure and compliant

Deploy into VPC and restrict access using security groups and IAM policies; support HIPAA, PCI, and ISO compliance

Cost effective

Pay-as-you-use pricing without any upfront costs or minimum requirements

Tightly integrated with other AWS services

Seamless data ingestion, security, auditing, and orchestration

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How does OpenSearch work

1

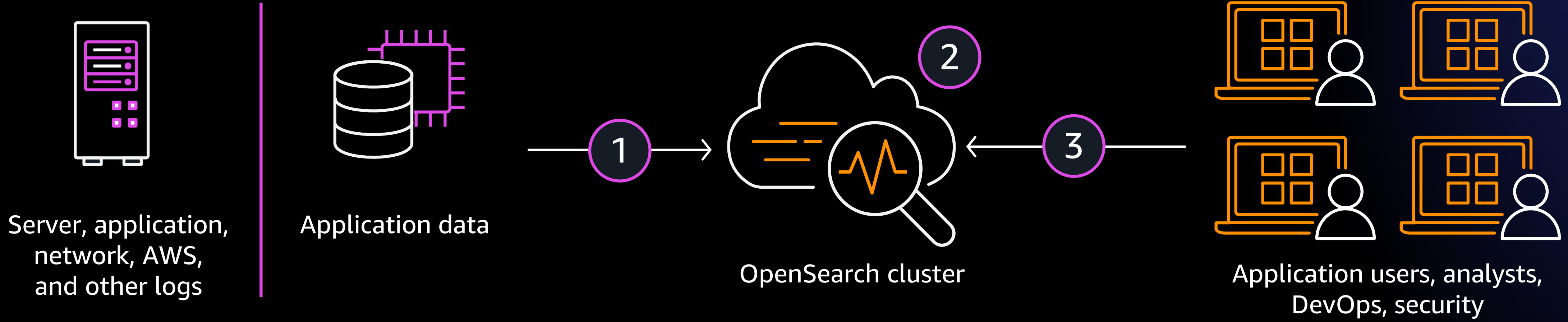
Send data as JSON via REST APIs

2

Data is indexed—all fields searchable, including nested JSON

3

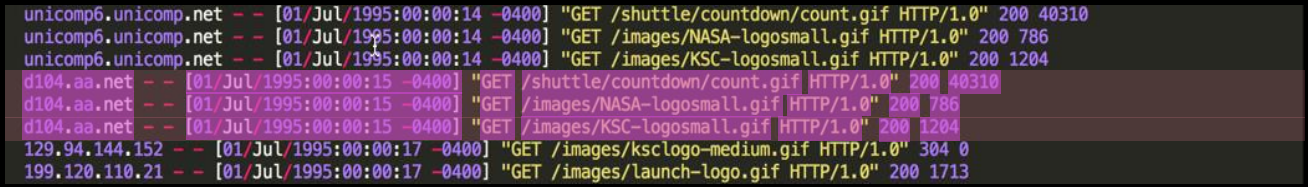
REST APIs, for fielded matching, Boolean expressions, sorting, and analysis



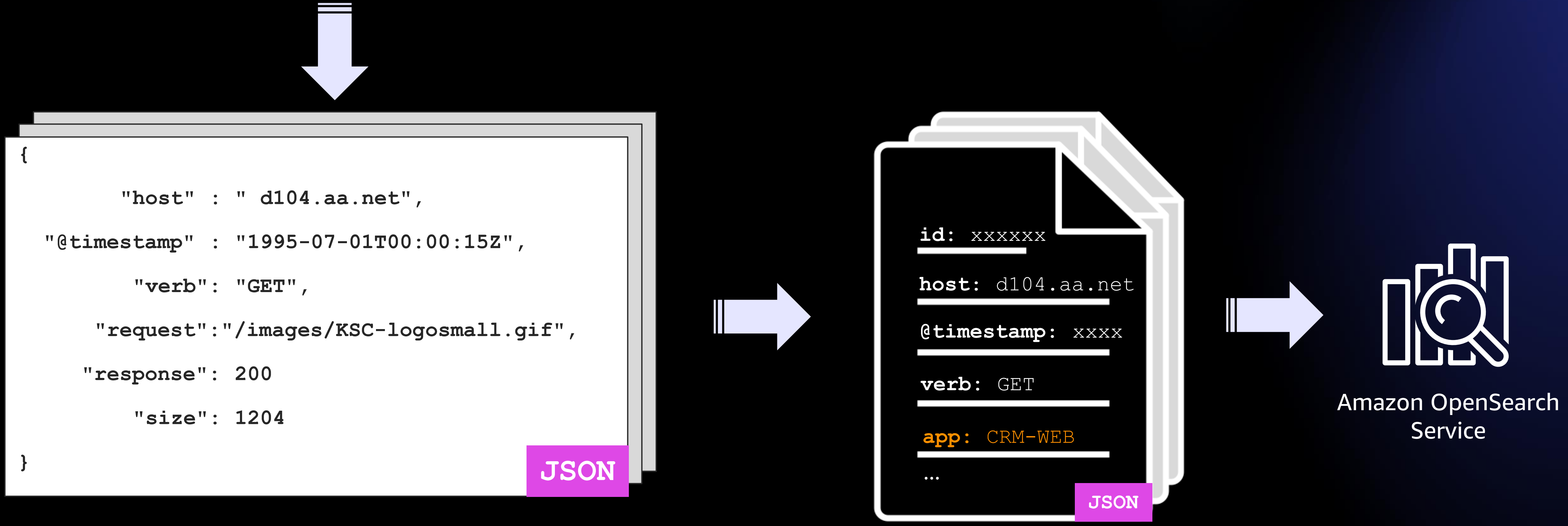
Server, application, network, AWS, and other logs
Application data
1
2
OpenSearch cluster
3
Application users, analysts, DevOps, security

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Storing logs as documents



```
unicomp6.unicomp.net - - [01/Jul/1995:00:00:14 -0400] "GET /shuttle/countdown/count.gif HTTP/1.0" 200 40310
unicomp6.unicomp.net - - [01/Jul/1995:00:00:14 -0400] "GET /images/NASA-logosmall.gif HTTP/1.0" 200 786
unicomp6.unicomp.net - - [01/Jul/1995:00:00:14 -0400] "GET /images/KSC-logosmall.gif HTTP/1.0" 200 1204
d104.aa.net - - [01/Jul/1995:00:00:15 -0400] "GET /shuttle/countdown/count.gif HTTP/1.0" 200 40310
d104.aa.net - - [01/Jul/1995:00:00:15 -0400] "GET /images/NASA-logosmall.gif HTTP/1.0" 200 786
d104.aa.net - - [01/Jul/1995:00:00:15 -0400] "GET /images/KSC-logosmall.gif HTTP/1.0" 200 1204
129.94.144.152 - - [01/Jul/1995:00:00:17 -0400] "GET /images/ksclogo-medium.gif HTTP/1.0" 304 0
199.120.110.21 - - [01/Jul/1995:00:00:17 -0400] "GET /images/launch-logo.gif HTTP/1.0" 200 1713
```



```
{
        "host" : " d104.aa.net",
  "@timestamp" : "1995-07-01T00:00:15Z",
         "verb": "GET",
      "request":"/images/KSC-logosmall.gif",
     "response": 200
         "size": 1204
}
```

id: xxxxxx

host: d104.aa.net

@timestamp: xxxx

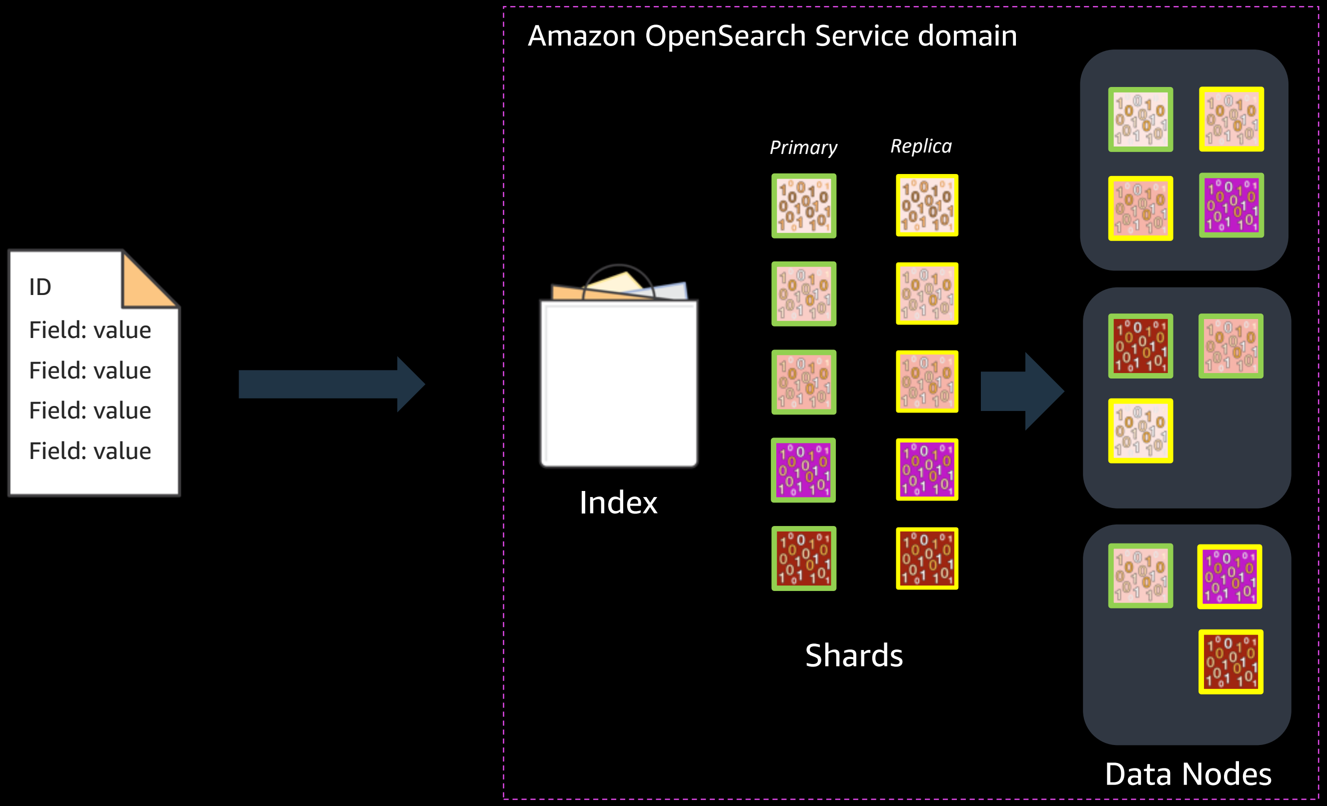
verb: GET

app: CRM-WEB

…

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

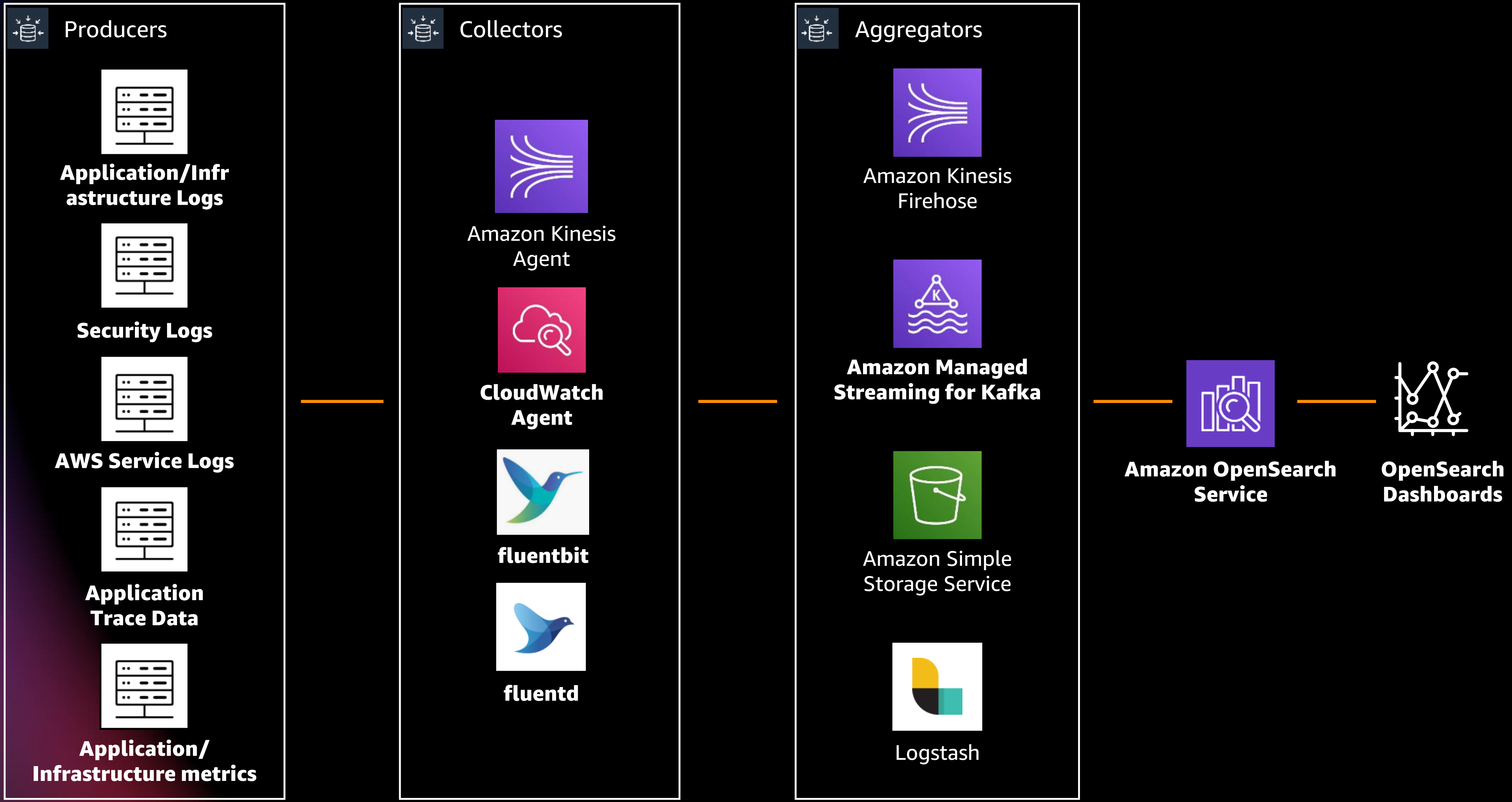
Data is stored in indexes, distributed across shards



- The index is an abstract entity it holds a *corpus* of documents
- Shards are distinct sets of documents. They store and compute
- OpenSearch distributes shards to data nodes
- Shards are primary or replica

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon OpenSearch Service data ingestion



Producers

Application/Infrastructure Logs

Security Logs

AWS Service Logs

Application Trace Data

Application/ Infrastructure metrics

Collectors

Amazon Kinesis Agent

CloudWatch Agent

fluentbit

fluentd

Aggregators

Amazon Kinesis Firehose

Amazon Managed Streaming for Kafka

Amazon Simple Storage Service

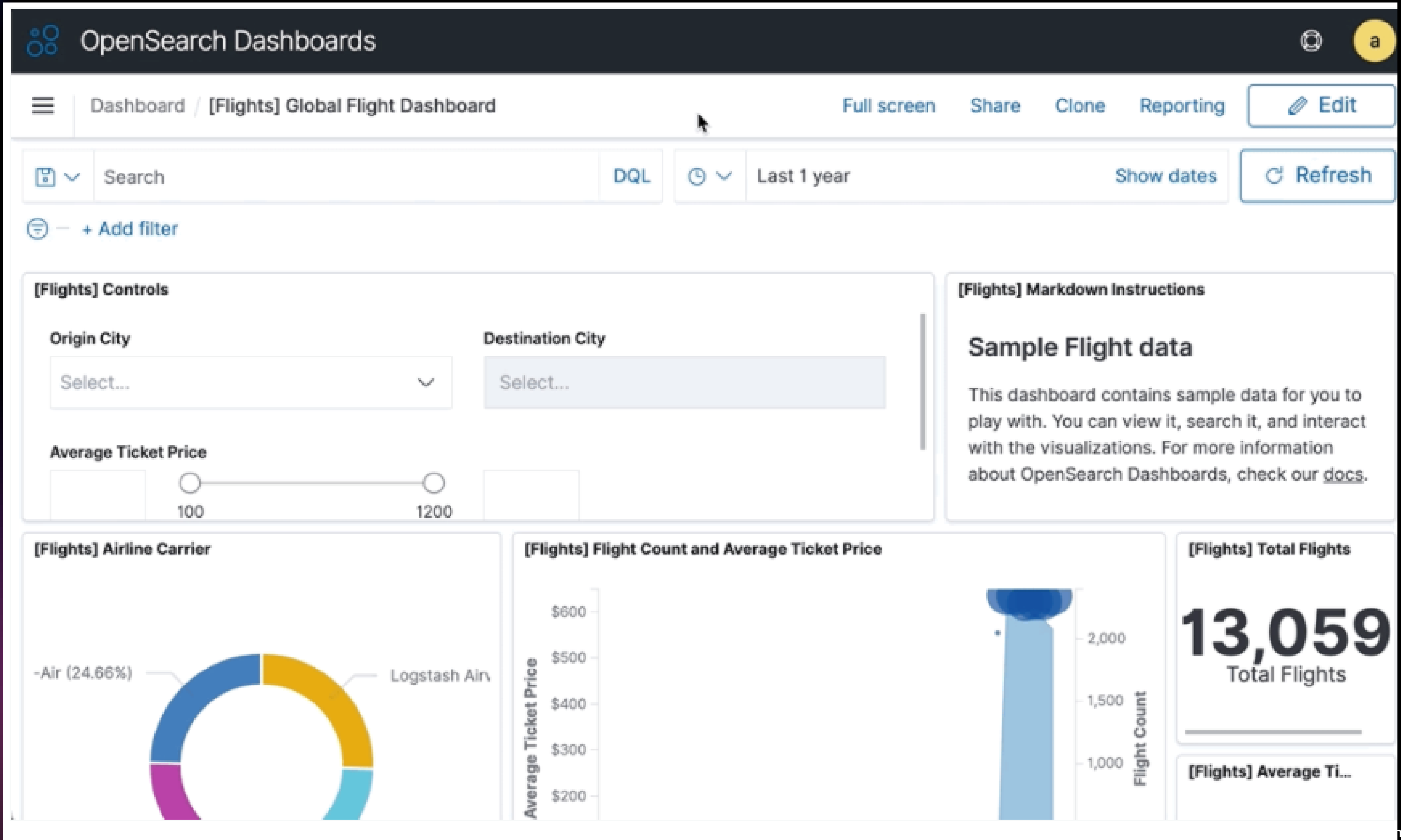
Logstash

Amazon OpenSearch Service

OpenSearch Dashboards

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Visualise data in OpenSearch dashboard

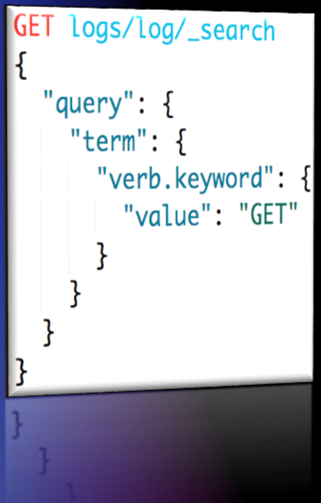


Sample Flight data

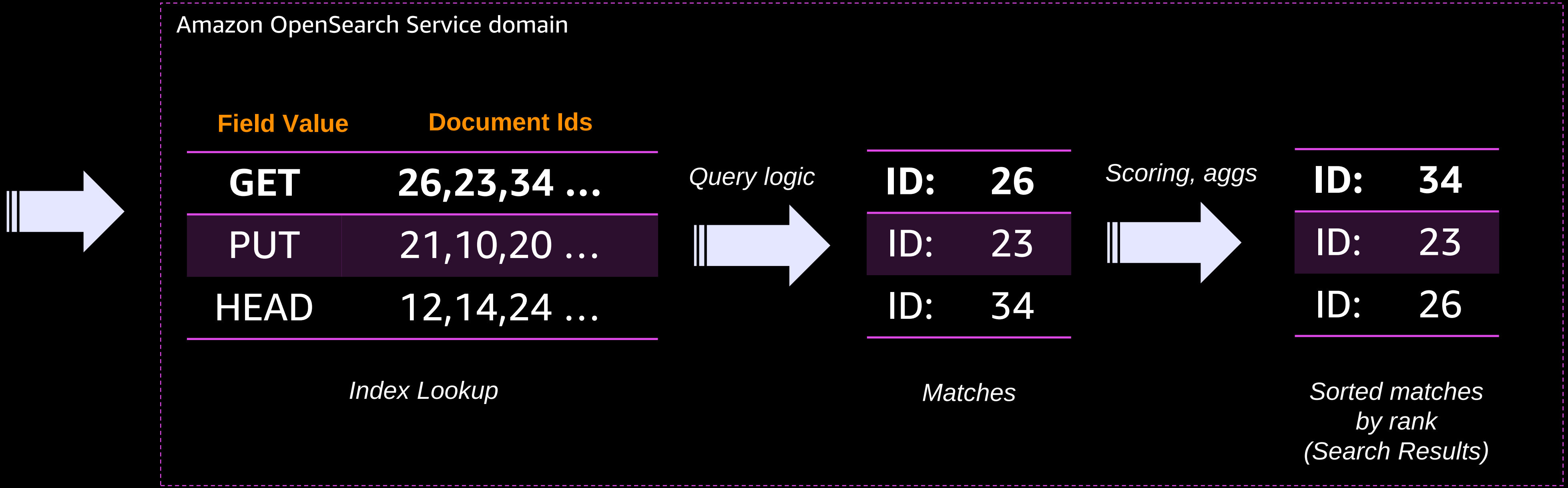
This dashboard contains sample data for you to play with. You can view it, search it, and interact with the visualizations. For more information about OpenSearch Dashboards, check our docs.

nc. or its affiliates. All rights reserved.

Use query API to retrieve data from Amazon OpenSearch Service



```
GET logs/log/_search
{
  "query": {
    "term": {
      "verb.keyword": {
        "value": "GET"
      }
    }
  }
}
```



Amazon OpenSearch Service domain

Field Value	Document Ids
GET	**26,23,34 …**
PUT	21,10,20 …
HEAD	12,14,24 …

Index Lookup

Query logic

ID: 26
ID: 23
ID: 34

Matches

Scoring, aggs

ID: 34
ID: 23
ID: 26

Sorted matches
by rank
(Search Results)

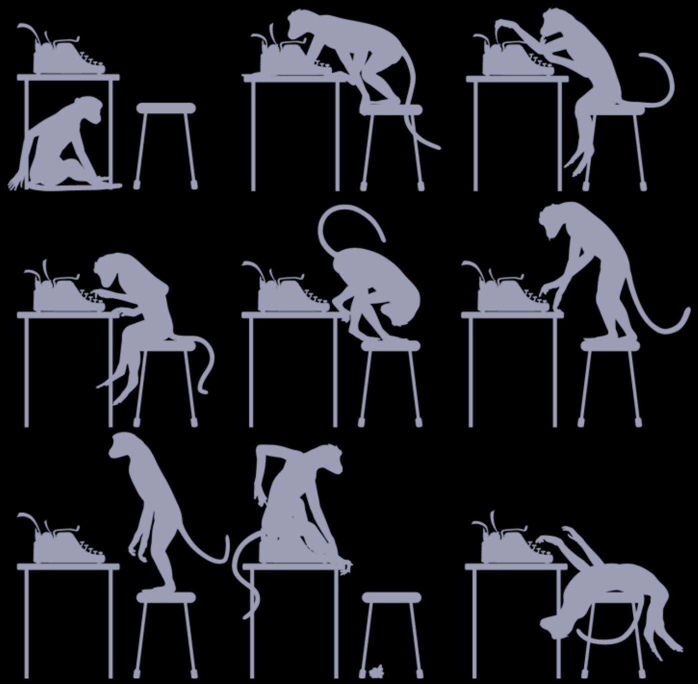
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Best practices

Get best practices for
Amazon OpenSearch Service

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Sharding strategy



- Set primary shard based on storage volume, recommended shard size between 30GB and 50GB (*test your shard sizes for optimal indexing/search throughput*)
- Always use at least 1 replica in production
- Set shard count in index template to achieve recommended shard size
- Review sharding strategy regularly to ensure you are staying close to recommended shard sizes

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Indexing naming and rotation

crm-web-2021-08-26

crm-web-2021-08-27

crm-app-2021-08-26

crm-app-2021-08-27

- Create index with root string (e.g. crm-web, crm-app) for easier index pattern creation for searching.
- Create index rotation frequency based on volume e.g. if you are receiving large volume then daily rotation.
- Daily index simplifies index management.
- Optimize rotation to achieve recommended shard size.
- Use aliases from start to avoid search clients configuration updates due to any naming/indexing strategy changes

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

There is no substitute to testing

- Benchmark your cluster search and indexing throughput.
- Test different sharding and indexing strategies to find optimal indexing and searching throughput
- Usually incorrect sharding strategy are responsible for cluster performance issues. Validate your sharding strategy by testing peak search traffic and data volumes
- Determine the limits of your cluster configuration and scaling thresholds by testing

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Create recommended Amazon CloudWatch alarms

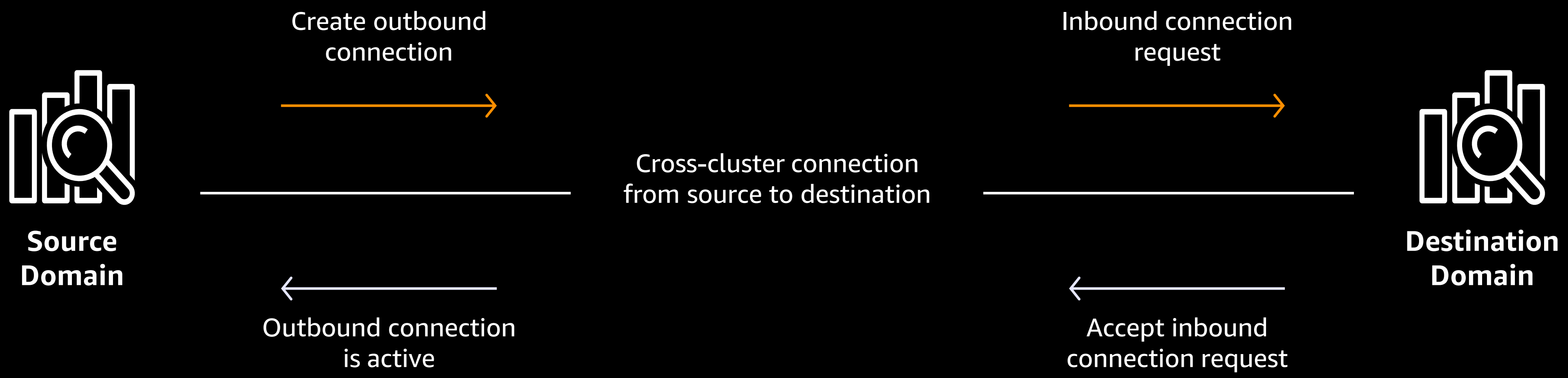
Name	Metric	Threshold	For Periods
ClusterStatus.red	Maximum	>= 1	1 x 1 min
ClusterIndexWritesBlocked	Maximum	>= 1	1 x 5 mins
CPUUtilization/MasterCPUUtilization	Average	>= 80%	3 x 15 mins
JVMMemoryPressure/Master…	Maximum	>= 80%	3 x 5 mins
FreeStorageSpace	Minimum	<= (25% of avail space)	1 x 1 min
AutomatedSnapshotFailure	Maximum	>= 1	1 x 1 min

https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/cloudwatch-alarms.html

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Cross-cluster search for Amazon OpenSearch Service

INCREASE SCALABILITY, EFFICIENCY & AVAILABILITY, BY SEPARATING DISTINCT WORKLOADS



Create outbound connection

Inbound connection request

Source Domain

Cross-cluster connection from source to destination

Destination Domain

Outbound connection is active

Accept inbound connection request

- Single OpenSearch dashboards interface to search across all included domains
- Tune domain resources for specific workloads
- Isolate failures to specific workloads

https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/cross-cluster-search.html

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

UltraWarm, low cost storage tier for Amazon OpenSearch Service

Store massive amounts of log data

Run interactive log analytics and visualization

Higher performance and durability

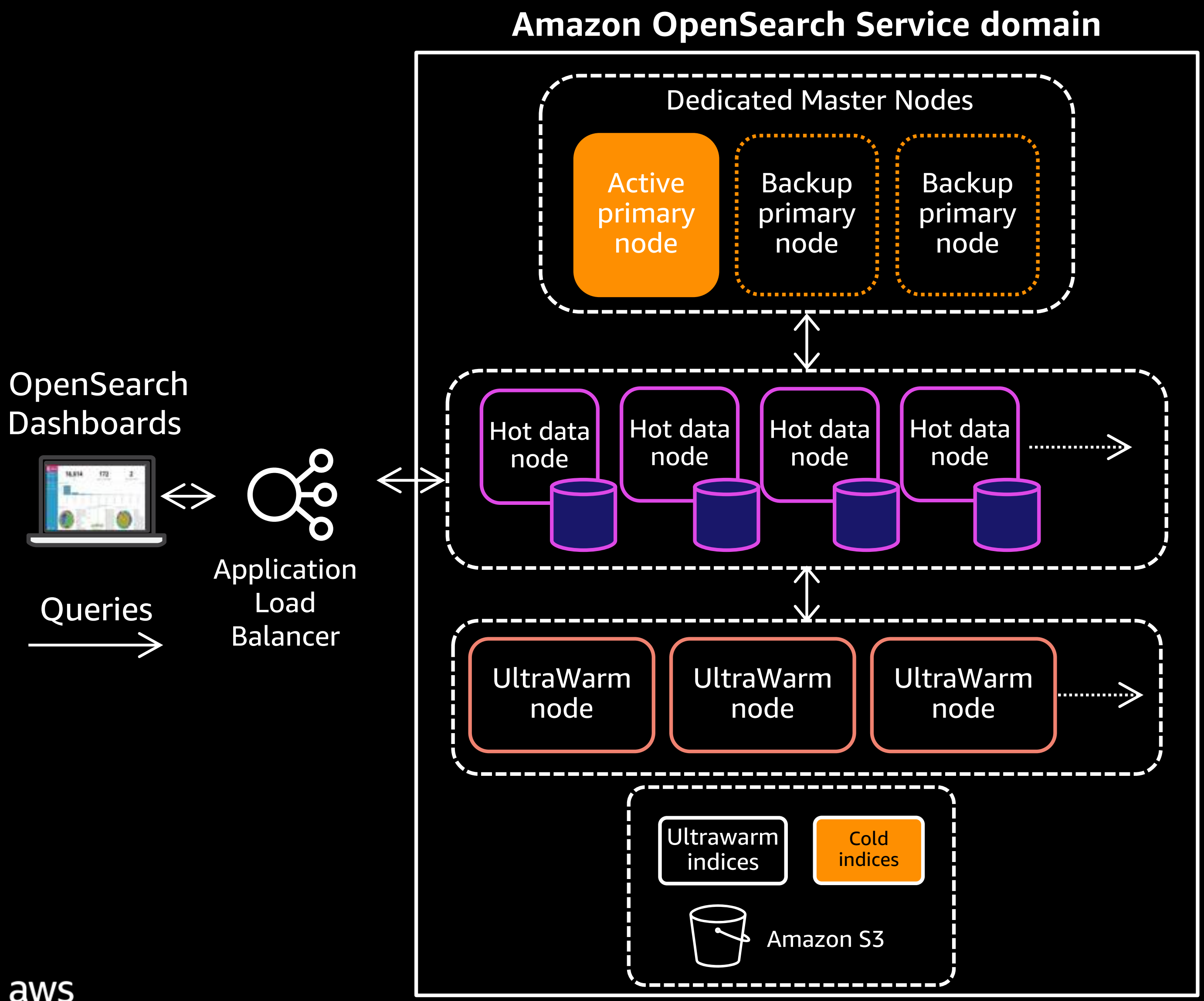
Achieve up to 90% cost savings

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

UltraWarm for Amazon OpenSearch Service

A WARM STORAGE TIER FOR AMAZON OPENSEARCH SERVICE

New Cold storage!



90% lower cost

Scale up to 3 PB per domain

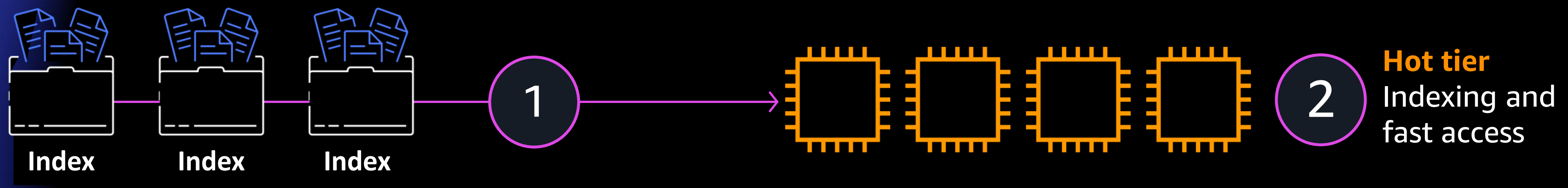
Analyze years of operational data

Interactive log analytics and visualization

https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/ultrawarm.html

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Index State Management (ISM) for data lifecycle

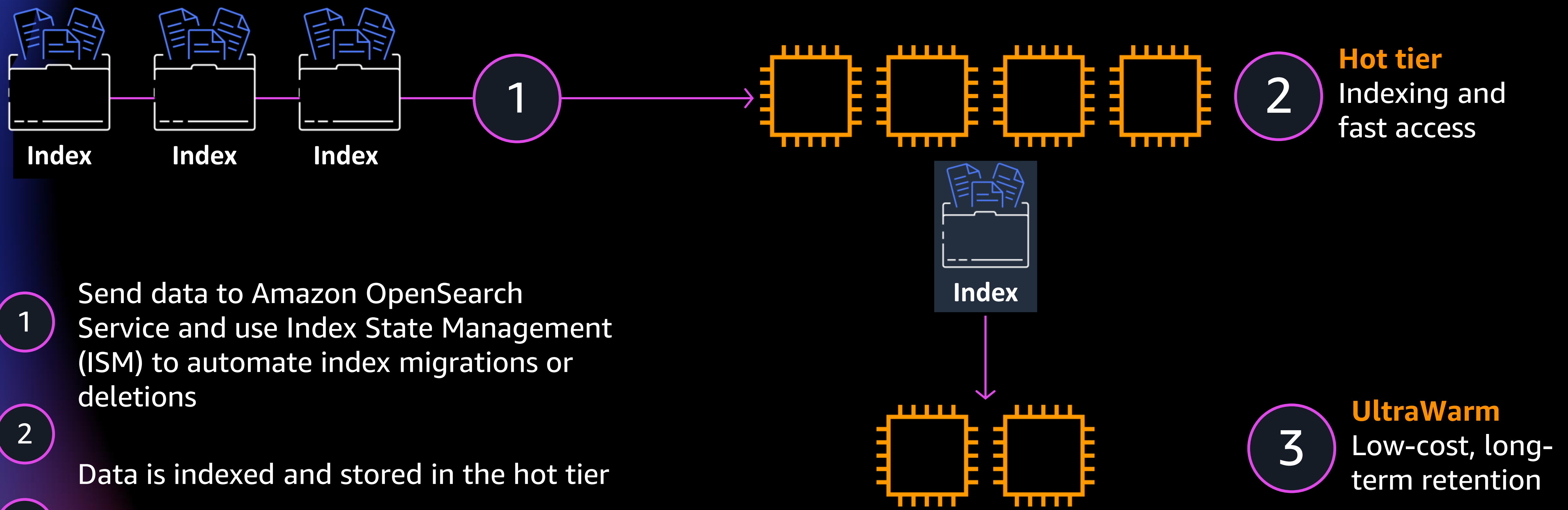


1. Send data to Amazon OpenSearch Service and use Index State Management (ISM) to automate index migrations or deletions

2. Data is indexed and stored in the hot tier

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Index State Management (ISM) for data lifecycle



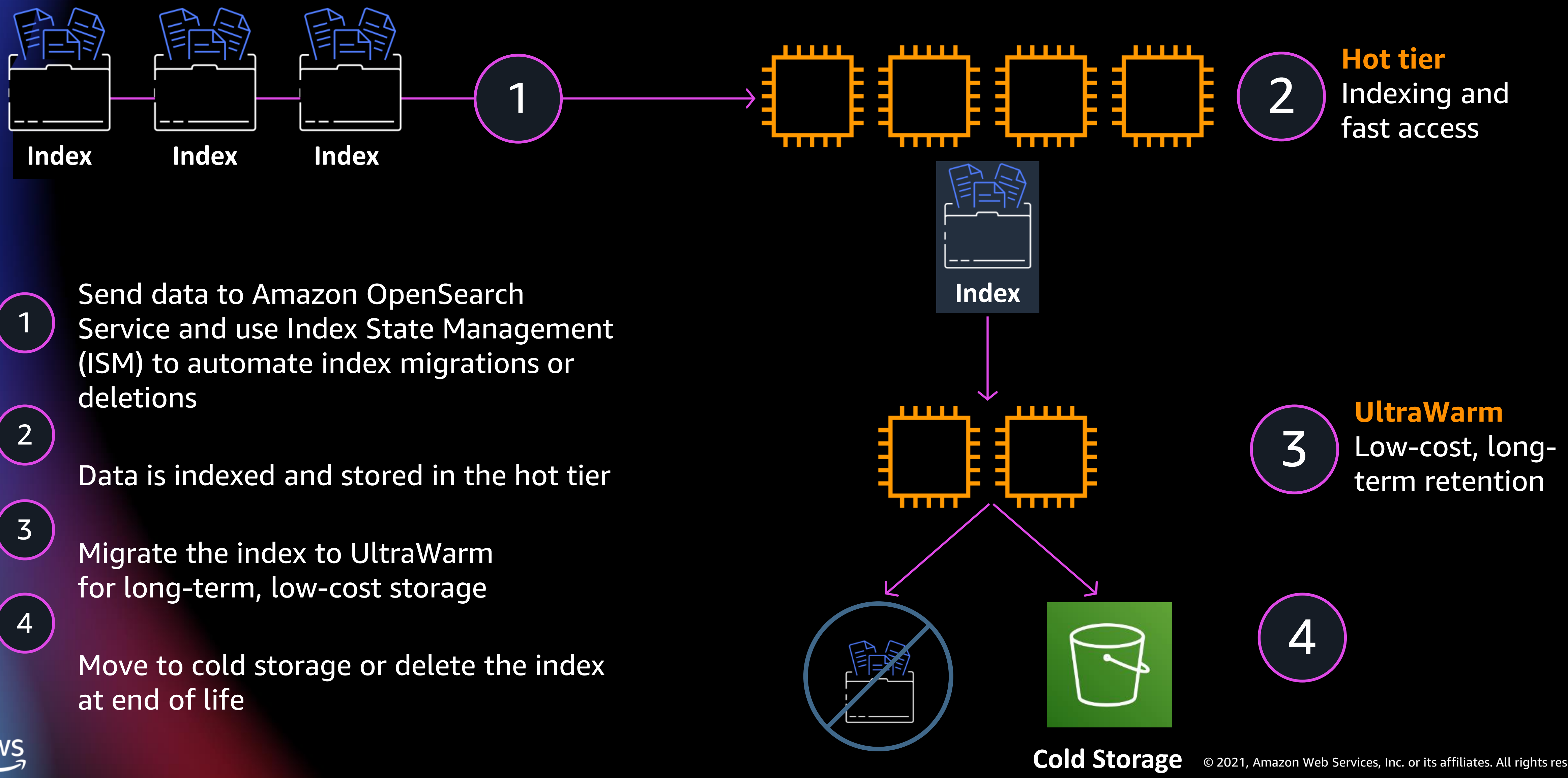
Hot tier
Indexing and fast access

UltraWarm
Low-cost, long-term retention

1. Send data to Amazon OpenSearch Service and use Index State Management (ISM) to automate index migrations or deletions
2. Data is indexed and stored in the hot tier
3. Migrate the index to UltraWarm for long-term, low-cost storage

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Index State Management (ISM) for data lifecycle



1 Send data to Amazon OpenSearch Service and use Index State Management (ISM) to automate index migrations or deletions

2 Data is indexed and stored in the hot tier

3 Migrate the index to UltraWarm for long-term, low-cost storage

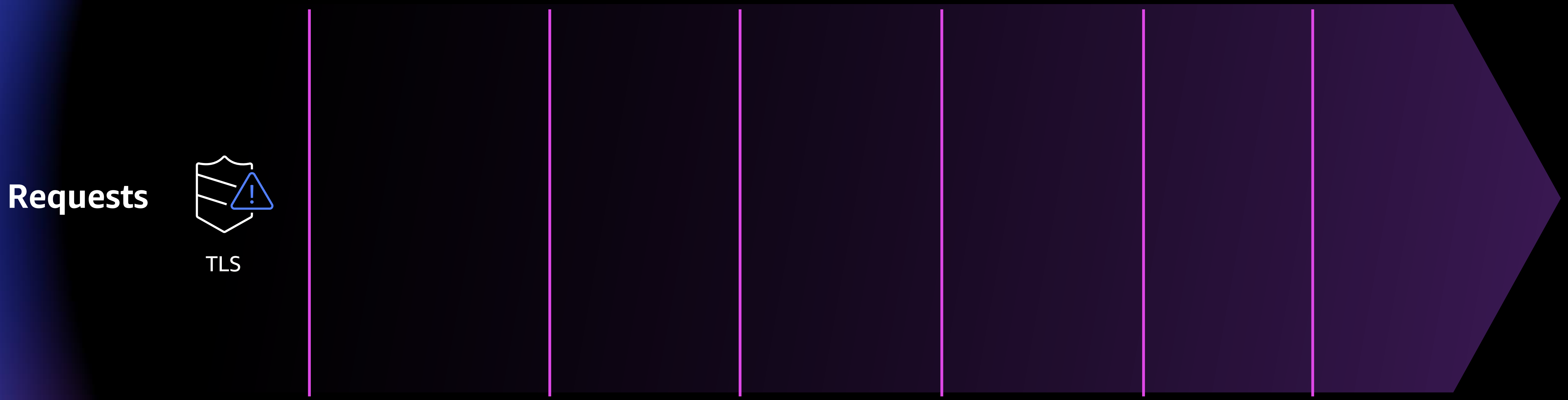
4 Move to cold storage or delete the index at end of life

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Security

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

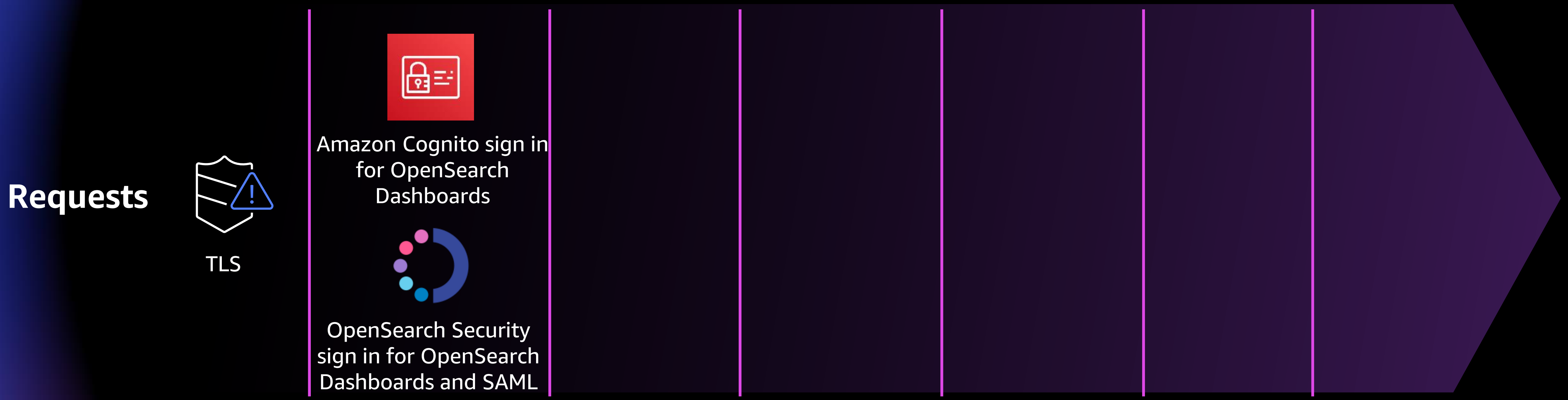
Multi-layer security with Amazon OpenSearch Service



- Encrypted from end to end—in flight with Transport Layer Security (TLS), at rest with AWS Key Management Service (KMS).
- Use a private endpoint to deploy into your Amazon Virtual Private Cloud (VPC) and security groups for traffic control.
- Includes OpenSearch Dashboards login via Amazon Cognito integration, or native with OpenSearch Security and SAML.
- Coarse-grained access control with AWS Identity and Access Management (IAM) policies.
- Fine-grained access control for tighter control over your data.

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Multi-layer security with Amazon OpenSearch Service



Requests

TLS

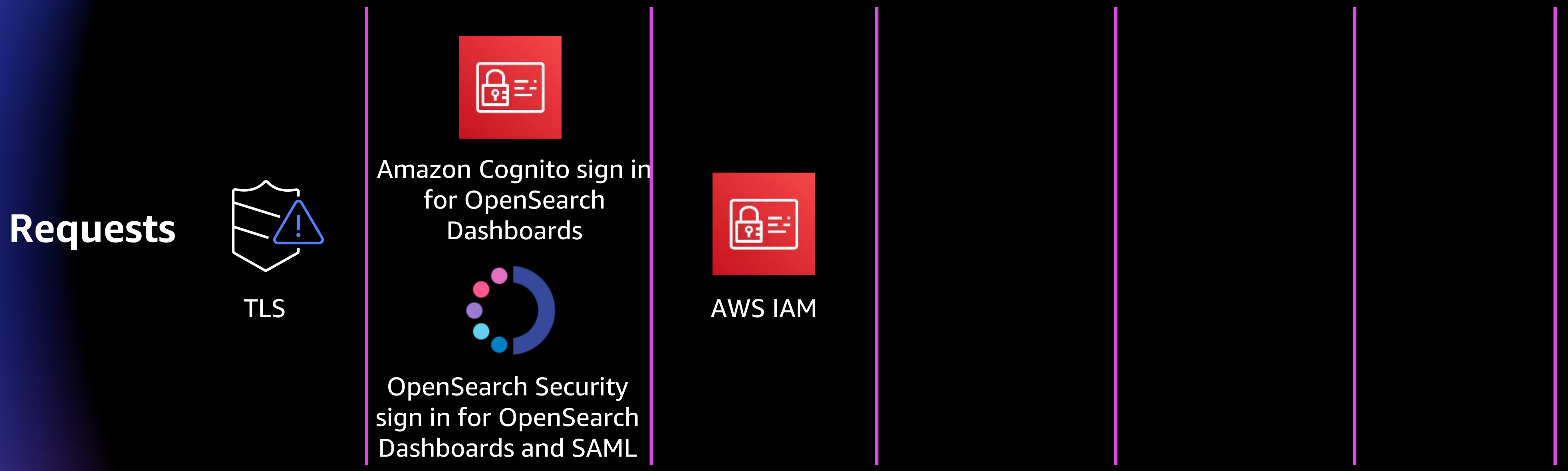
Amazon Cognito sign in for OpenSearch Dashboards

OpenSearch Security sign in for OpenSearch Dashboards and SAML

- Encrypted from end to end—in flight with Transport Layer Security (TLS), at rest with AWS Key Management Service (KMS).
- Use a private endpoint to deploy into your Amazon Virtual Private Cloud (VPC) and security groups for traffic control.
- Includes OpenSearch Dashboards login via Amazon Cognito integration, or native with OpenSearch Security and SAML.
- Coarse-grained access control with AWS Identity and Access Management (IAM) policies.
- Fine-grained access control for tighter control over your data.

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Multi-layer security with Amazon OpenSearch Service



Requests

TLS

Amazon Cognito sign in for OpenSearch Dashboards

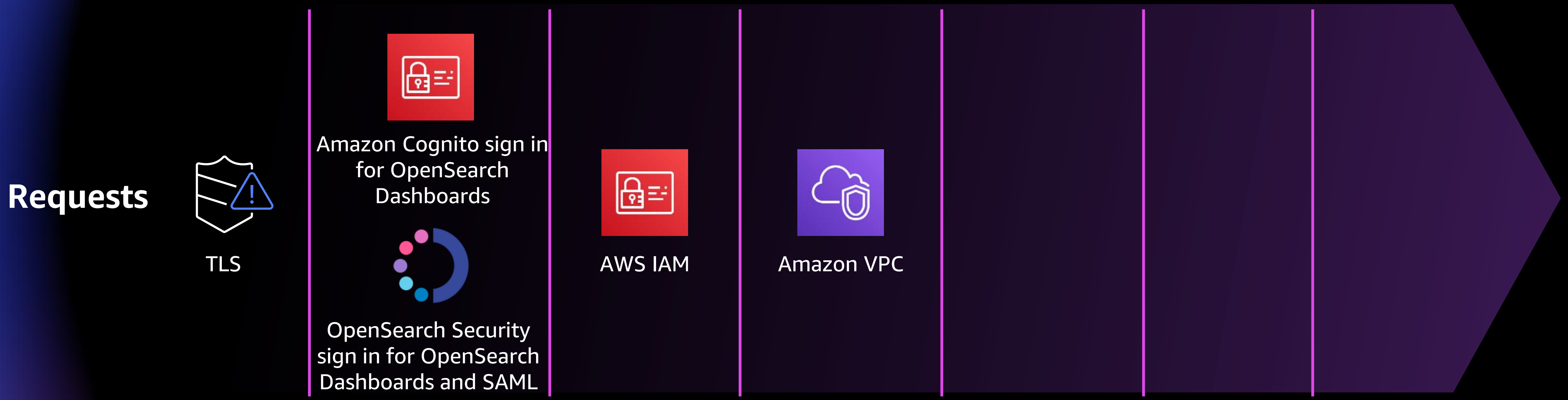
OpenSearch Security sign in for OpenSearch Dashboards and SAML

AWS IAM

- Encrypted from end to end—in flight with Transport Layer Security (TLS), at rest with AWS Key Management Service (KMS).
- Use a private endpoint to deploy into your Amazon Virtual Private Cloud (VPC) and security groups for traffic control.
- Includes OpenSearch Dashboards login via Amazon Cognito integration, or native with OpenSearch Security and SAML.
- Coarse-grained access control with AWS Identity and Access Management (IAM) policies.
- Fine-grained access control for tighter control over your data.

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

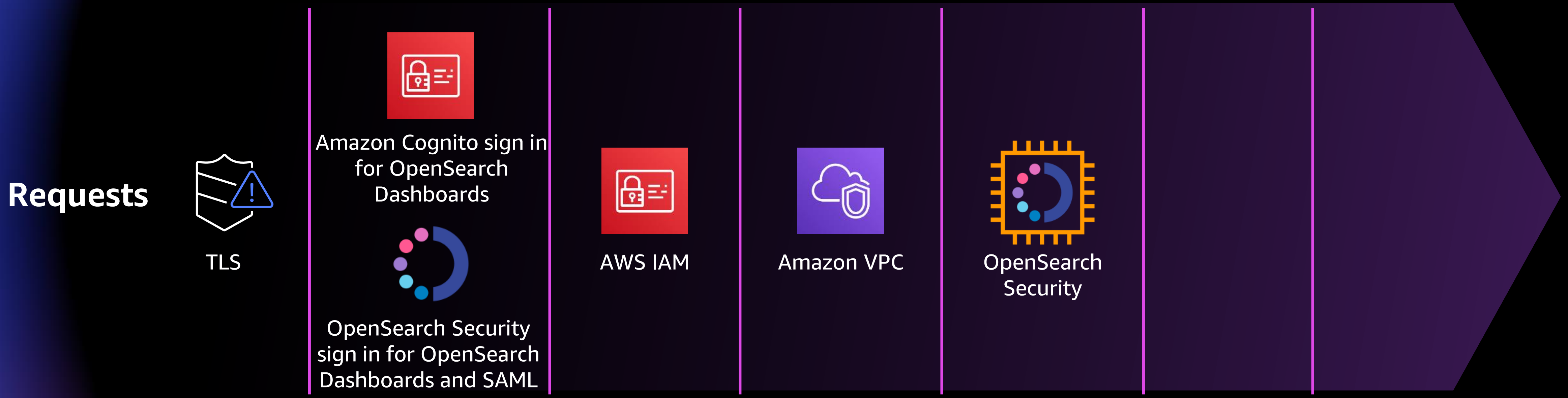
Multi-layer security with Amazon OpenSearch Service


Requests
TLS
Amazon Cognito sign in for OpenSearch Dashboards
OpenSearch Security sign in for OpenSearch Dashboards and SAML
AWS IAM
Amazon VPC

- Encrypted from end to end—in flight with Transport Layer Security (TLS), at rest with AWS Key Management Service (KMS).
- Use a private endpoint to deploy into your Amazon Virtual Private Cloud (VPC) and security groups for traffic control.
- Includes OpenSearch Dashboards login via Amazon Cognito integration, or native with OpenSearch Security and SAML.
- Coarse-grained access control with AWS Identity and Access Management (IAM) policies.
- Fine-grained access control for tighter control over your data.

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

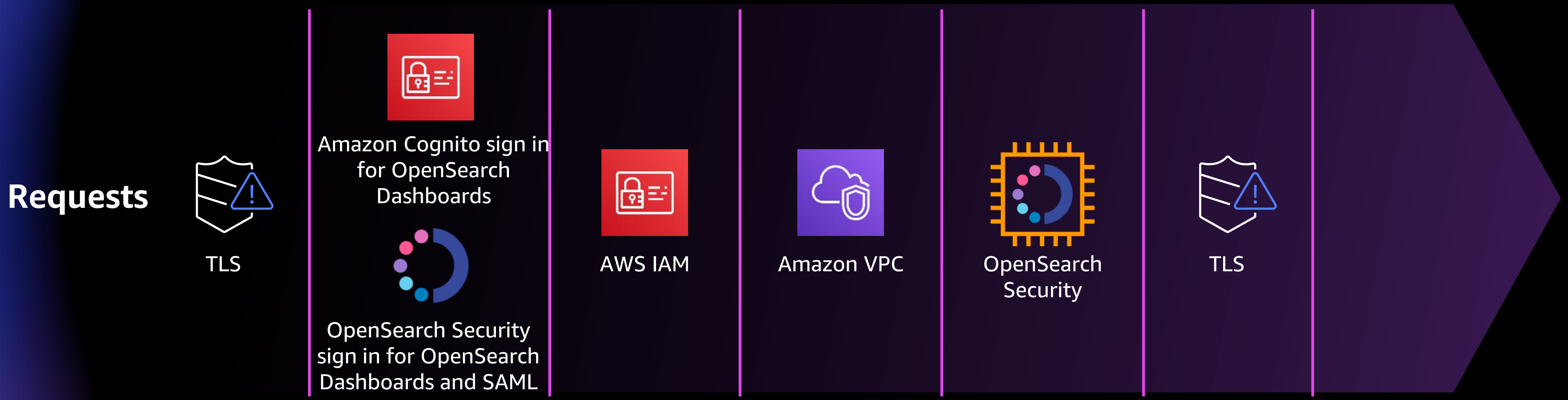
Multi-layer security with Amazon OpenSearch Service


Requests
TLS
Amazon Cognito sign in for OpenSearch Dashboards
OpenSearch Security sign in for OpenSearch Dashboards and SAML
AWS IAM
Amazon VPC
OpenSearch Security

- Encrypted from end to end—in flight with Transport Layer Security (TLS), at rest with AWS Key Management Service (KMS).
- Use a private endpoint to deploy into your Amazon Virtual Private Cloud (VPC) and security groups for traffic control.
- Includes OpenSearch Dashboards login via Amazon Cognito integration, or native with OpenSearch Security and SAML.
- Coarse-grained access control with AWS Identity and Access Management (IAM) policies.
- Fine-grained access control for tighter control over your data.

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

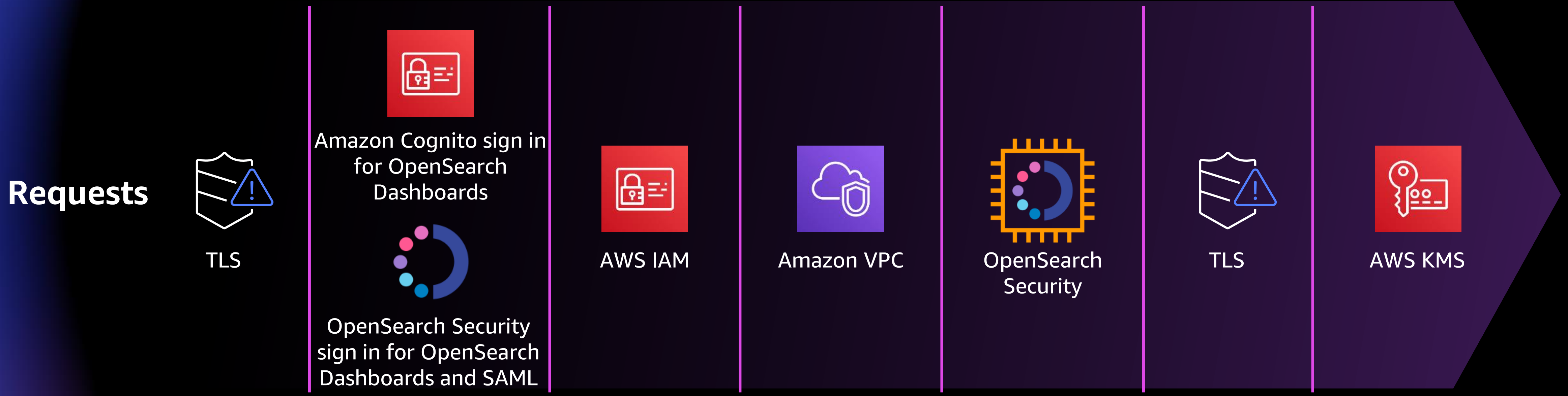
Multi-layer security with Amazon OpenSearch Service


Requests
TLS
Amazon Cognito sign in for OpenSearch Dashboards
OpenSearch Security sign in for OpenSearch Dashboards and SAML
AWS IAM
Amazon VPC
OpenSearch Security
TLS

- Encrypted from end to end—in flight with Transport Layer Security (TLS), at rest with AWS Key Management Service (KMS).
- Use a private endpoint to deploy into your Amazon Virtual Private Cloud (VPC) and security groups for traffic control.
- Includes OpenSearch Dashboards login via Amazon Cognito integration, or native with OpenSearch Security and SAML.
- Coarse-grained access control with AWS Identity and Access Management (IAM) policies.
- Fine-grained access control for tighter control over your data.

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Multi-layer security with Amazon OpenSearch Service


Requests
TLS
Amazon Cognito sign in for OpenSearch Dashboards
OpenSearch Security sign in for OpenSearch Dashboards and SAML
AWS IAM
Amazon VPC
OpenSearch Security
TLS
AWS KMS

- Encrypted from end to end—in flight with Transport Layer Security (TLS), at rest with AWS Key Management Service (KMS).
- Use a private endpoint to deploy into your Amazon Virtual Private Cloud (VPC) and security groups for traffic control.
- Includes OpenSearch Dashboards login via Amazon Cognito integration, or native with OpenSearch Security and SAML.
- Coarse-grained access control with AWS Identity and Access Management (IAM) policies.
- Fine-grained access control for tighter control over your data.

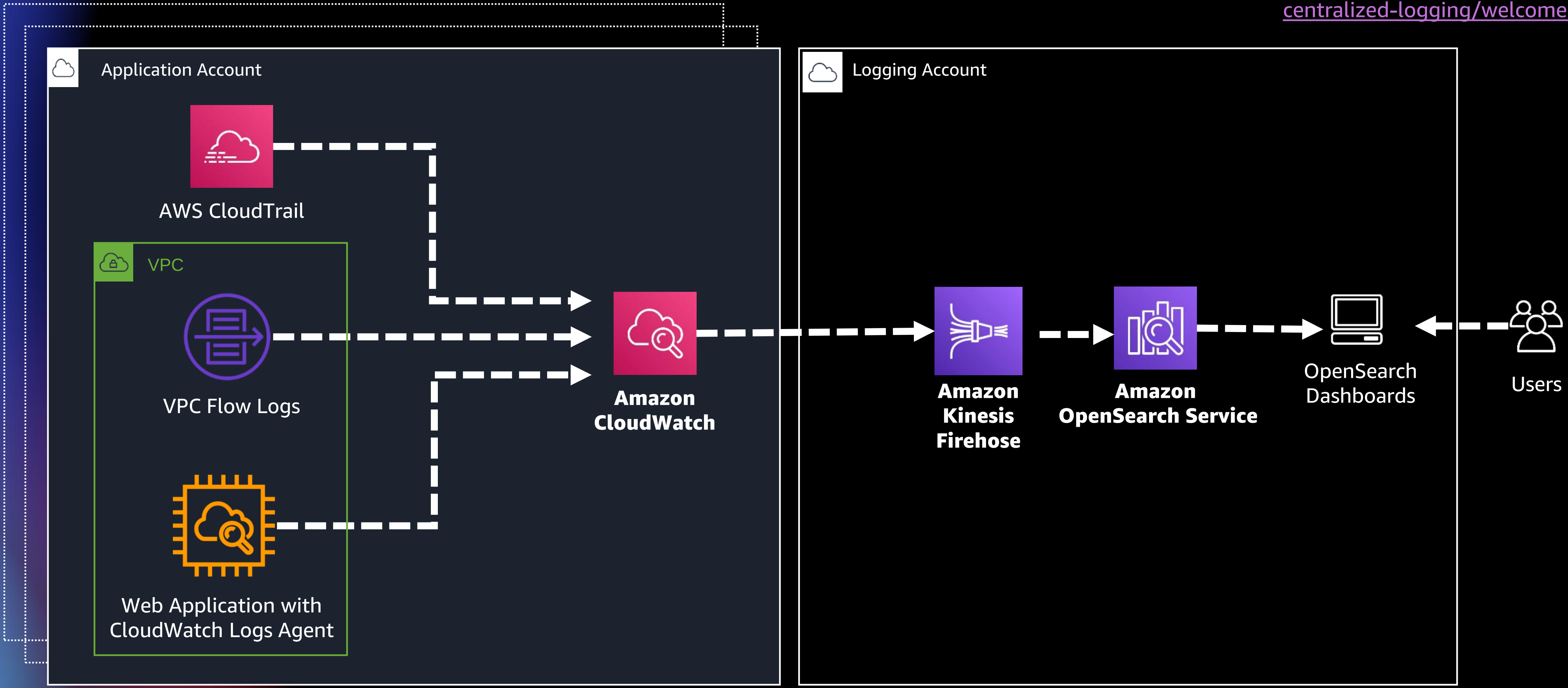
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Demo

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Multi-Account log analytics architecture demo

https://docs.aws.amazon.com/solutions/latest/centralized-logging/welcome.html



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Recap

- Analyzing your machine generated data can give you valuable insights that can create efficiencies and differentiate your business
- Amazon OpenSearch Service is purpose built service for log analytics
- There are few key best practices that you can use to get the best out of your OpenSearch clusters
- Using Ultra-warm, cross cluster searching you can scale your workloads
- OpenSearch provide you fine grain access control so you can give your user access to only the data that they own

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Additional resources

- https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/es-gsg.html
- https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/cloudwatch-alarms.html
- https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/aes-bp.html
- https://aws.amazon.com/blogs/big-data/introducing-cold-storage-for-amazon-elasticsearch-service/
- https://docs.aws.amazon.com/solutions/latest/centralized-logging/welcome.html

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Visit the AWS Data Resource Hub

Dive deeper with these resources, get inspired and learn how you can use data to make better decisions and innovate faster.

- Building a winning data strategy
- The new leadership mindset for data & analytics
- Harness data to reinvent your organization
- Put your data to work with a modern analytics approach
- Breaking free from on-premises database constraints
- Cloud storage adoption: From cost optimization to agility & innovation
- A strategic playbook for data, analytics, and machine learning
- … and more!

https://tinyurl.com/aws-data-resource

Visit resource hub

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Training and Certification

Empower your teams with comprehensive training

By building skills with AWS Training and Certification, businesses and individuals can see the bigger picture understanding the reasoning behind every data point. As training progresses and teams become data-fluent, previously hidden insights come into view.

Leverage free digital training

Learn how to harness the world's most valuable resource: data. Access digital and virtual instructor-led courses on data analytics and databases built by the experts at AWS and start your learning journey to become data-driven.

Take a digital course »

Get certified

Earn industry-recognized credibility and set tangible goals for success with industry-recognized certifications, like *AWS Certified Data Analytics – Specialty*.

Learn more »

Ramp-up your skills

Deep dive into new topics and focus on knowledge gaps at your own pace with the *AWS Ramp-Up Guide: Database* and *AWS Ramp-Up Guide: Data Analytics*. With a wide range of whitepapers, blog posts, videos, webinars and peer resources available for data professionals to leverage for independent learning.

Download ramp-up guides »

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you for attending AWS Innovate – Data Edition

We hope you found it interesting! A kind reminder to **complete the survey.**
Let us know what you thought of today's event and how we can improve the event experience for you in the future.

aws-apj-marketing@amazon.com

twitter.com/AWSCloud

facebook.com/AmazonWebServices

youtube.com/user/AmazonWebServices

slideshare.net/AmazonWebServices

twitch.tv/aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.